PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04N 5/21 — A1

(11) International Publication Number: WO 99/62249

(43) International Publication Date: 2 December 1999 (02.12.99)

(21) International Application Number: PCT/US99/12040

(22) International Filing Date: 28 May 1999 (28.05.99)

(30) Priority Data:
09/087,736 29 May 1998 (29.05.98) US

(71) Applicant: CONEXANT SYSTEMS, INC. [US/US]; 4311 Jamboree Road, Newport Beach, CA 92660 (US).

(72) Inventor: DIMEO, Matthew, J.; 4075 Vermont Street, San Diego, CA 92103 (US).

(74) Agent: LAND, John; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US).

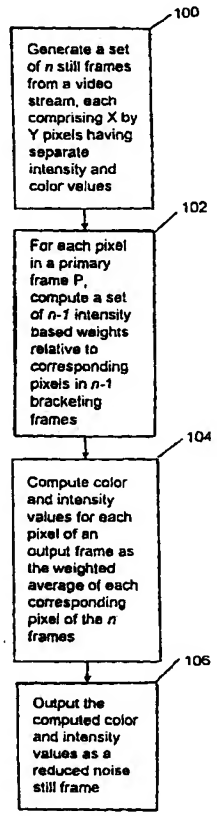
(81) Designated States: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD AND SYSTEM FOR COMBINING MULTIPLE IMAGES INTO A SINGLE HIGHER–QUALITY IMAGE

(57) Abstract

A method for combining a set of n still image frames into a single reduced noise still image. The invention reduces visable noise, along with some television color artifacts, in still images captured from a video stream by combining data from multiple video frames. Each frame comprises an XY array of pixel data (100). In the preferred embodiment, each pixel includes data defining an intensity level as well as color values (100). A weighted average of several frames of video is taken to generate a single still output frame (102, 104). The weighting is determined adaptively in order to reduce interframe motion artifacts (102, 104). The resulting output has substantially less visable noise than a simple unweighted average (106). The algorithm is much faster and simpler to implement than true multi–pixel motion estimation, which is not really necessary in a series of nominally–still images.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PNSDOCID: <WO 9962249A1 I >

METHOD AND SYSTEM FOR COMBINING MULTIPLE IMAGES INTO A SINGLE HIGHER-QUALITY IMAGE

TECHNICAL FIELD

This invention relates to electronic video capture systems, and more particularly to a method and system for combining multiple image frames into a single reduced noise still image.

BACKGROUND

"Video still capture" involves a process of extracting a still image from an image set comprising a sequence of electronically-encoded image frames. A number of video capture products are on the market that allow a user to capture one or more still images from a conventional video source to an attached computer, such as a personal computer (PC). The video source can be a signal conforming to one of the well-known NTSC, PAL, or SECAM video standards, or a video camera, or sequential images from a digital camera, or recorded video signals played on a video cassette recorder (VCR) or the like.

Many video sources, such as consumer-quality computer cameras and consumer VCRs, tend to have noise problems that degrade image quality. When capturing single video frames for display as still images, these problems are still visible. Some attempts have been made to reduce such noise. For example, the SNAPPY™ video snapshot product from Play, Inc. generates an unweighted average of multiple video fields without any motion compensation to output a single still image with somewhat reduced noise.

The inventor has determined that there is still a need for improved noise reduction in a video capture system. The present invention provides a method and system for achieving this end.

SUMMARY

The invention reduces visible noise, along with some television color artifacts, by combining a set of *n* still image frames into a single reduced noise still image. Each frame comprises an XY array of picture element (pixel) data. In the preferred embodiment, each pixel includes data defining an intensity level (Y in the YUV color space model) as well as color values (UV in the YUV color space model). However, other color space models may be used if the models define each pixel as having color space coordinates defining intensity and color values, directly or indirectly (*i.e.*, by having separate intensity and color values, or where such information can be derived from the color space coordinates, or where the color space coordinates can be separately weighted). A weighted average of several frames of video is taken to generate a single still output frame. The weighting is determined adaptively in order to reduce inter-frame motion artifacts.

The resulting output has substantially less visible noise than a simple unweighted average. The algorithm described above is much faster and simpler to implement than true multi-pixel motion estimation, which is not really necessary in a series of nominally-still images. If averaging is done on individual NTSC-derived images in the YUV or YIQ color spaces, and if weights are based only on pixel intensity, television-related color artifacts may also be reduced by a comb filtering effect.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 is a block diagram of a prior art electronic video tuner system and computer system that may be used with the present invention.

FIG. 2 is a flowchart of a preferred embodiment of the invention.

Like reference numbers and designations in the various drawings indicate like elements.

RNSDOCID: <WO 9962249A1 | >

DETAILED DESCRIPTION

The invention will be described in the context of a conventional image capture system the provides a series of decoded video image frames to a buffer memory of a computer system. The invention operates on the contents of the buffer memory to provide a still image with reduced noise.

FIG. 1 is a block diagram of a prior art electronic video decoder system 1 and computer system 2 that may be used with the present invention. Any of several video sources, such as a video camera 3, VCR 4, or a television antenna or cable signal source 5 processed through an RF tuner 5a, are coupled to the video decoder system 1, which may be, for example, a Rockwell Semiconductor Systems Bt848A single-chip video capture processor and PCI Bus Master, available from Rockwell Semiconductor Systems, Newport Beach, CA. The video decoder system 1 can place video data directly into the system memory 6 of the computer system 2 for video capture applications. The computer system 2 includes the system memory 6, a central processing unit 7, and a host bridge 8 to couple the computer system to a conventional bus 9, such as the PCI Bus. The video decoder system 1 can also provide video data to a computer graphics adapter 10, which includes a graphics controller 11, a frame buffer 12, and a RAMDAC 13. The output of the video decoder system 1 is placed into a target area in the frame buffer 12 for video overlay applications, so that a video image stream may be displayed on a monitor 14.

The video decoder system 1 includes an analog-to-digital converter (ADC) 20 for converting the input video signal to a "raw" digital format. A Y/C separation circuit 21 separates luminance and color (chroma) information encoded within the digital signal into conventional YUV color space data values. The separated information is temporarily stored in a first-in-first-out (FIFO) memory 22. A PCI Bus Master circuit 23 provides controlled transfer of decoded, digitized video frames over the bus 9 connecting the video decoder system 1 to the computer system 2. (Although the illustrated embodiment shows use of a PCI bus, the invention is not limited to any particular bus). A video synchronization and scaling circuit 24 locks the video decoder system 1 to the incoming analog video signal and allows resizing of the video image size. Control of the video decoder system 1 is provided through either a general purpose input/output port 25 or a two-wire inter-integrated circuit interface

26. In an alternative embodiment, Y/C separation may be performed within the computer system 2 using appropriate programming.

For purposes of the invention, the video decoder system 1 is used to place data for a series of still video frames 6-1 ... 6-*n* directly into the system memory 6 of the computer system 2, as shown in FIG. 1. If the data is not in YUV format, Y/C separation is performed by appropriate processing of such data to generate YUV values for each pixel of the series of still video frames 6-1 ... 6-*n*. One method of separating luminance (Y) from chrominance (C) in a video signal is simply to use a notch filter to generate Y, and a bandpass filter to extract C. The U and V components of the YUV color space are derived from C by quadrature amplitude demodulation, in known fashion. Thereafter, in the preferred embodiment, a computer implementation of the method described below is applied to a set of *n* video frames to derive a still image with reduced visible noise. The computer implementation preferably executes on the CPU 7 of the computer system 2.

The invention reduces visible noise, along with some television color artifacts, in still images captured from a video stream by combining data from multiple video frames. Each frame comprises an XY array of picture element (pixel) data. In the preferred embodiment, each pixel includes data defining an intensity level (Y in the YUV color space model) as well as color values (UV in the YUV color space model). A weighted average of several frames of video is taken to generate a single still output frame. The weighting is determined adaptively in order to reduce inter-frame motion artifacts.

More particularly, one frame near the middle (in time) of a sequence of *n* frames is designated the "primary" frame *P*. Its weight, *Wp*, is "1" for every pixel. The intensity level Y*p* of each pixel of the primary frame is compared to the intensity level of the corresponding pixel of each of the other *n*-1 bracketing frames. Based on the absolute difference between the Y value of a primary-frame pixel and that of another frame *n*, a weight *Wn* is generated. Thus, every pixel location has its own set of *n* weights. Each weight *Wn* is set to one if the difference is below a selected threshold T1 and to zero if the difference is above a selected threshold T2. The values for T1 and T2 may be empirically determined. Between the two thresholds, the weight varies from one to zero, preferably linearly (although other functions may be used). To summarize, each weight *W* is calculated as follows:

RNSDOCID: <WO 9962249A1 | >

If $|Yp - Yn| < T1$, $Wn = 1$

Else If $|Yp - Yn| > T2$, $Wn = 0$

Else $Wn = (|Yp - Yn| - T2) / (T1-T2)$

Once weights have been determined, YUV values for each output pixel of the desired still frame are calculated as follows:

$$Yout = (W_1*Y_1 + \ldots + W_n*Y_n) / (W_1 + \ldots + W_n)$$

$$Uout = (W_1*U_1 + \ldots + W_n*U_n) / (W_1 + \ldots + W_n)$$

$$Vout = (W_1*V_1 + \ldots + W_n*V_n) / (W_1 + \ldots + W_n)$$

where $(Y_1, U_1, V_1) \ldots (Y_n, U_n, V_n)$ are the corresponding pixel values from each frame 1 to n.

The resulting output has substantially less visible noise than a simple unweighted average. The algorithm described above is much faster and simpler to implement than true multi-pixel motion estimation, which is not really necessary in a series of nominally-still image. Since averaging is done in the YUV color space, and since weights are based only on intensity rather than partially on color information, television-related color artifacts are reduced.

While the YUV color space model is shown in the illustrated embodiment, other color space models may be used if the models define each pixel as having color space coordinates defining intensity and color values, directly or indirectly. For example, a suitable color space model may have separate intensity and color values, or such information can be derived from the color space coordinates, or the color space coordinates can be separately weighted. For instance, for an RGB color space model, there is no separate intensity channel. However, each color axis (R, G, B) can be weighted separately, or the RGB values for a pixel can be used to calculate back to an intensity value which is used for weighting in accordance with the invention. Other color space models to which the invention can be adapted are YCrCb and YIQ.

FIG. 2 is a flowchart of a preferred embodiment of the invention. Initially, a set of still frames are generated by the video decoder system 1 and placed in the system memory 6

of the computer system 2 (STEP 100). Each frame comprises X by Y pixels, each pixel having color space coordinates defining intensity and color values, directly or indirectly (*e.g.*, YUV) values. Next, for each pixel in a primary frame *P*, a set of *n* intensity-based weights relative to corresponding pixels in *n*-1 bracketing frames are computed as described above (STEP 102). Color and intensity values are computed for each pixel of an output frame as the weighted average of each corresponding pixel of the *n* frames, as described above (STEP 104). Lastly, the computed color and intensity values of the output frame are output (*e.g.*, to a monitor or data file) as a reduced noise still frame (STEP 106).

Computer Implementation

Aspects of the invention may be implemented in hardware or software, or a combination of both. However, preferably, the algorithms and processes of the invention are implemented in one or more computer programs executing on programmable computers each comprising at least one processor, at least one data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. Program code is applied to input data to perform the functions described herein and generate output information. The output information is applied to one or more output devices, in known fashion.

Each program may be implemented in any desired computer language (including machine, assembly, high level procedural, or object oriented programming languages) to communicate with a computer system. In any case, the language may be a compiled or interpreted language.

Each such computer program is preferably stored on a storage media or device (*e.g.*, ROM, CD-ROM, tape, or magnetic diskette) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The inventive system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

A number of embodiments of the present invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. For example, the captured still frame may be directly output to a printer, or transmitted electronically to another computer. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for combining a set of *n* still image frames into a single reduced noise still image, each frame comprising X by Y pixels, each pixel having color space coordinates defining intensity and color values, including the steps of:
   (a) for each pixel in a primary frame *P* of the set of still image frames, computing a set of *n* intensity-based weights W*n* relative to corresponding pixels in *n*-1 bracketing still image frames;
   (b) computing color and intensity values for each pixel of an output frame as the weighted average of each corresponding pixel of the *n* still image frames;
   (c) outputting the computed output frame as a reduced noise still frame.

2. The method of claim 1, wherein each pixel in each still image frame *n* has an intensity value Y*n*, wherein each pixel in the primary frame *P* has an intensity value Y*p*, and the step of computing a set of *n* intensity-based weights W*n* includes the step of computing each weight W*n* according to the relationships:

   If $|Yp - Yn| < T1$, $Wn = 1$

   Else If $|Yp - Yn| > T2$, $Wn = 0$

   Else $Wn = (|Yp - Yn| - T2) / (T1-T2)$,

   where T1 and T2 are selectable thresholds.

3. The method of claim 2, wherein each pixel in each still image frame n has color values Un and Vn, and wherein the step of computing color and intensity values for each pixel of an output frame includes the step of computing YUV values for each pixel of the output frame according to the relationships:

$$Yout = (W_1 * Y_1 + \ldots + W_n * Y_n) / (W_1 + \ldots + W_n)$$

$$Uout = (W_1 * U_1 + \ldots + W_n * U_n) / (W_1 + \ldots + W_n)$$

$$Vout = (W_1 * V_1 + \ldots + W_n * V_n) / (W_1 + \ldots + W_n),$$

where $(Y_1, U_1, V_1) \ldots (Y_n, U_n, V_n)$ are the corresponding pixel intensity and color values from each still image frame 1 to n.

4. A system for combining a sequence of electronically-encoded image frames into a single reduced noise still image, including:

(a) a video decoder system and computer system for generating a set of *n* still image frames, each still image frame comprising X by Y pixels, each pixel having color space coordinates defining intensity and color values; and

(b) a programmed processor in the computer system for:

(1) computing a set of *n* intensity-based weights W*n* relative to corresponding pixels in *n*-1 bracketing still image frames for each pixel in a primary frame *P* of the set of still frames; and

(2) computing color and intensity values for each pixel of an output frame as the weighted average of each corresponding pixel of the *n* still image frames, the output frame having reduced noise.

5. The system of claim 4, further including an output device, coupled to the computer system, for outputting the computed output frame as a reduced noise still frame.

6. The system of claim 4, wherein each pixel in each still image frame *n* has an intensity value Y*n*, wherein each pixel in the primary frame *P* has an intensity value Y*p*, and the program for computing a set of *n* intensity-based weights W*n* further computes each weight W*n* according to the relationships:

If $|Yp - Yn| < T1$, $Wn = 1$

Else If $|Yp - Yn| > T2$, $Wn = 0$

Else $Wn = (|Yp - Yn| - T2) / (T1-T2)$,

where T1 and T2 are selectable thresholds.

BNSDOCID: <WO 9962249A1 >

7. The system of claim 6, wherein each pixel in each still image frame *n* has color values U*n* and V*n*, and wherein the program for computing color and intensity values for each pixel of an output frame further computes YUV values for each pixel of the output frame according to the relationships:

$$Yout = (W_1 * Y_1 + \ldots + W_n * Y_n) / (W_1 + \ldots + W_n)$$

$$Uout = (W_1 * U_1 + \ldots + W_n * U_n) / (W_1 + \ldots + W_n)$$

$$Vout = (W_1 * V_1 + \ldots + W_n * V_n) / (W_1 + \ldots + W_n),$$

where $(Y_1, U_1, V_1) \ldots (Y_n, U_n, V_n)$ are the corresponding pixel intensity and color values from each still image frame 1 to *n*.

8. A computer program, residing on a computer-readable medium, for combining a set of *n* still image frames into a single reduced noise still image, each frame comprising X by Y pixels, each pixel having color space coordinates defining intensity and color values, the computer program comprising instructions for causing a computer to:

(a) for each pixel in a primary frame *P* of the set of still image frames, compute a set of *n* intensity-based weights W*n* relative to corresponding pixels in *n*-1 bracketing still image frames;

(b) compute color and intensity values for each pixel of an output frame as the weighted average of each corresponding pixel of the *n* still image frames;

(c) output the computed output frame as a reduced noise still frame.

9. The computer program of claim 8, wherein each pixel in each still image frame *n* has an intensity value Y*n*, wherein each pixel in the primary frame *P* has an intensity value Y*p*, and the instructions for causing a computer to compute a set of *n* intensity-based weights W*n* further includes instructions for causing a computer to compute each weight W*n* according to the relationships:

If $|Yp - Yn| < T1$, $Wn = 1$

Else If $|Yp - Yn| > T2$, $Wn = 0$

Else $Wn = (|Yp - Yn| - T2) / (T1-T2)$,

where T1 and T2 are selectable thresholds.

10. The computer program of claim 9, wherein each pixel in each still image frame *n* has color values U*n* and V*n*, and wherein the instructions for causing a computer to compute color and intensity values for each pixel of an output frame further includes instructions for causing a computer to compute YUV values for each pixel of the output frame according to the relationships:

$$Yout = (W_1 * Y_1 + \ldots + W_n * Y_n) / (W_1 + \ldots + W_n)$$

$$Uout = (W_1 * U_1 + \ldots + W_n * U_n) / (W_1 + \ldots + W_n)$$

$$Vout = (W_1 * V_1 + \ldots + W_n * V_n) / (W_1 + \ldots + W_n),$$

where $(Y_1, U_1, V_1) \ldots (Y_n, U_n, V_n)$ are the corresponding pixel intensity and color values from each still image frame 1 to *n*.

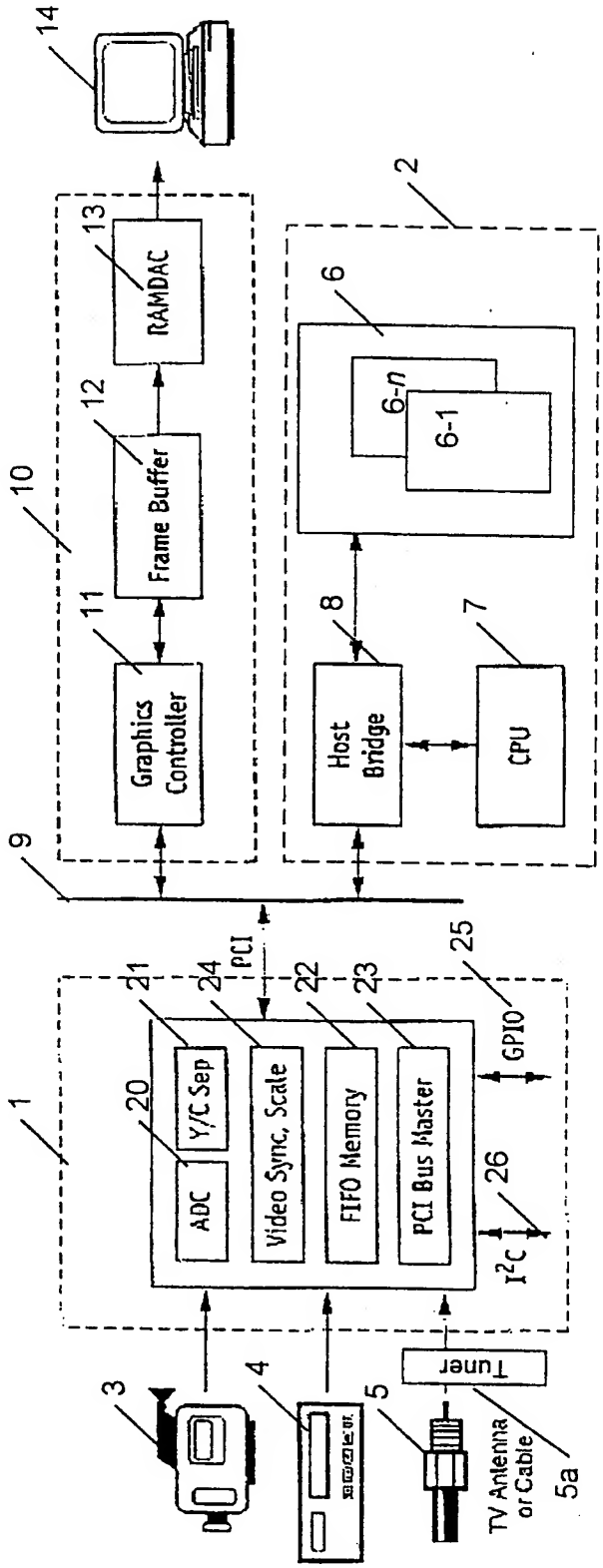


FIG. 1
(Prior Art)

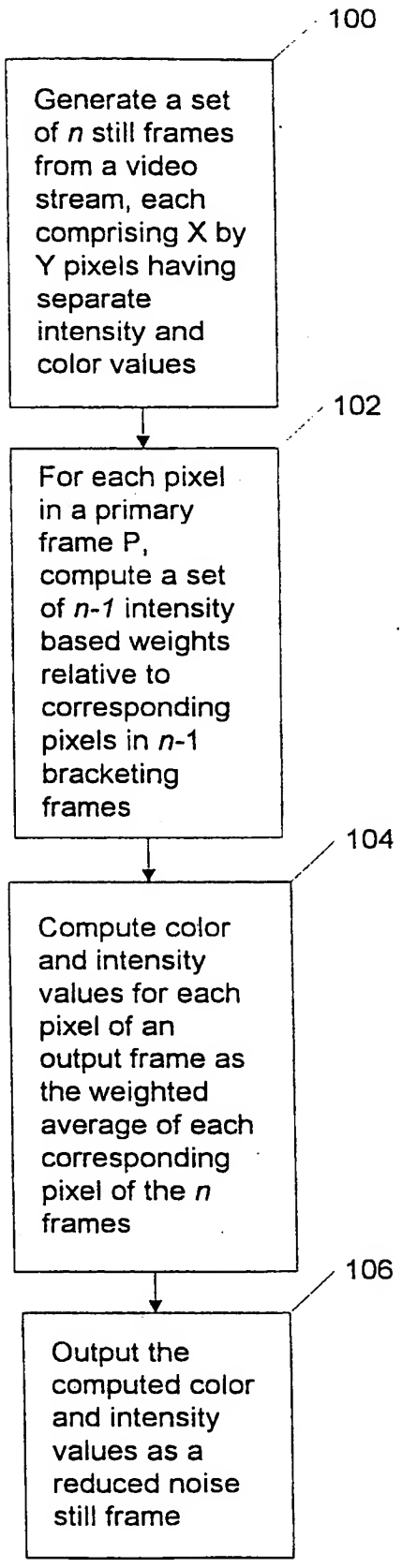

100
Generate a set of n still frames from a video stream, each comprising X by Y pixels having separate intensity and color values
102
For each pixel in a primary frame P, compute a set of n-1 intensity based weights relative to corresponding pixels in n-1 bracketing frames
104
Compute color and intensity values for each pixel of an output frame as the weighted average of each corresponding pixel of the n frames
106
Output the computed color and intensity values as a reduced noise still frame

FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12040

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :H04N 5/21
US CL :348/620
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U S 348/620, 618, 619, 624, 559, 560; 358/167,36; H04N/ 5/21

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

none

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4,987,481 A (SPEARS et al.) 22 January 1991, see entire document.	1,4,5,and 8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
19 JULY 1999	18 AUG 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer James R. Matthews DAVID E. HARVEY Telephone No. (703) 305-4365

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO 9962249A1 >